Supplement File

Clinical Trial Generalizability Assessment in the Big Data Era: A Review

Zhe He[1], Xiang Tang[2], Xi Yang[3], Yi Guo[3], Thomas J. George[4], Neil Charness[5], Kelsa Bartley Quan Hem[6], William Hogan[3], Jiang Bian[3]

[1]School of Information, Florida State University, Tallahassee, Florida, USA
[2]Department of Statistics, Florida State University, Tallahassee, Florida, USA
[3]Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, Florida, USA
[4]Hematology & Oncology, Department of Medicine, College of Medicine, University of Florida, Gainesville, Florida, USA
[5]Department of Psychology, Florida State University, Tallahassee, Florida, USA
[6]Calder Memorial Library, Miller School of Medicine, University of Miami, Miami, Florida, USA

Table S1. Search strategies used for the database searches.

Database / Search Engine	**Search Queries**	**Number of Returned Records**
MEDLINE / PubMed	(trial[Title/Abstract] OR clinical study[Title/Abstract] OR clinical trial[Title/Abstract]) AND (population representativeness[Title/Abstract] OR restrictive eligibility criteria[Title/Abstract] OR external validity[Title/Abstract] OR generalizability[Title/Abstract])	1921
CINAHL	(TI (trial OR "clinical trial" OR "clinical study") OR AB (trial OR "clinical trial" OR "clinical study")) AND (TI ("population representativeness" OR "restrictive eligibility criteria" OR "external validity" OR "generalizability") OR AB ("population representativeness" OR "restrictive eligibility criteria" OR "external validity" OR "generalizability"))	744
PychINFO	(TI (trial OR "clinical trial" OR "clinical study") OR AB (trial OR "clinical trial" OR "clinical study")) AND (TI ("population representativeness" OR "restrictive eligibility criteria" OR "external validity" OR "generalizability") OR AB ("population representativeness" OR "restrictive eligibility criteria" OR "external validity" OR "generalizability"))	1007
Cochrane	Title/Abstract/Keywords("clinical trial" OR "clinical study" OR "trial" OR "trial") AND Title/Abstract/Keywords("population representativeness" OR "restrictive eligibility criteria" OR "external validity" OR "generalizability")	1680

Table S2. Compared patient information (demographic information, clinical characteristics, adverse events, and outcomes) in *a posteriori* and *a priori* generalizability assessment papers.

Total (N=144)	**Demographic Information**		**Clinical Characteristics**		**Adverse Events**		**Outcomes**	
	N	Y	N	Y	N	Y	N	Y
A posteriori	2	118	23	97	114	6	65	55
A priori	6	35	10	31	40	1	26	15
Odds Ratio (p-value)	0.099 ** (0.006)		0.735 (0.475)		0.475 (0.497)		0.682 (0.304)	
95% CI	0.0191 - 0.5118		0.3156 - 1.7119		0.0555 - 4.0675		0.3286 - 1.4147	

* means p-value < 0.05, ** means p-value <0.01, *** means p-value < 0.001

Table S3. Examples of *a priori* and *a posteriori* generalizability assessment papers.

Type	Result Format	Methods	Reference
A priori	Score-based	GIST 2.0: "*We propose a multi-trait metric - GIST 2.0 that can compute the a priori generalizability based on the population representativeness of a clinical study by explicitly modeling the dependencies among all eligibility criteria."*	Sen[27]
	Score-based	mGIST: "*We extended a published metric named Generalizability Index for Study Traits (GIST) to include multiple study traits for quantifying the population representativeness of a set of related studies by assuming the independence and equal importance among all study traits."*	He[26]
	Non-score-based	*"We applied a standard set of eligibility criteria representative of GAD pharmacological and psychotherapy clinical trials to all adults with past 12 months GAD (n = 894), and to a subgroup of participants seeking treatment (n = 329). Our aim was to assess how many participants with GAD would fulfill typical eligibility criteria."*	Hoertel[29]
A posteriori	Score-based	*"We then propose a framework for a standardized evaluation of parameters relevant to determining the external validity of clinical trials to produce a 'generalizability score'. We then apply this framework to populations of patients with heart failure included in trials, cohorts and registries to demonstrate the use of the generalizability score and its graphic representation along three dimensions: participants' demographics, their clinical profile and intervention setting. We use the generalizability score to compare a single trial to multiple 'target' clinical scenarios. Additionally, we present the generalizability score of several studies with regard to a single "target" population."*	Cahan[34]
	Score-based	*"We propose the use of propensity-score-based metrics to quantify the similarity of the participants in a randomized trial and a target population. In this setting the propensity score model predicts participation in the randomized trial, given a set of covariates. The resulting propensity scores are used first to quantify the difference between the trial participants and the target population, and then to match, subclassify, or weight the control group outcomes to the population, assessing how well the propensity score-adjusted outcomes track the outcomes actually observed in the*	Stuart[35]

		population."	
	Non-score-based	*"The sample of the web-based SUD intervention (Therapeutic Education System vs. Treatment-as-usual; n = 507) was compared with the target population of SUD treatment-seeking individuals from the Treatment Episodes Data Set-Admissions (TEDS-A). Using weights based on the probabilities of RCT participation, we computed weighted treatment effects on retention and abstinence."*	de Jonghe[6]
	Non-score-based	*"The demographic, clinical and laboratory characteristics of HIV-infected participants in two antiretroviral trials (Concorde and Delta) at three study sites were compared with those of two other groups of patients to whom the trial results would be applicable: eligible patients who were screened for the trials but who did not enrol, and eligible patients who were not approached or screened for the trials."*	Moore[33]
A posteriori	Post hoc generalization	*Increasingly, the statistical and epidemiologic literature is focusing beyond issues of internal validity and turning its attention to questions of external validity. Here, we discuss some of the challenges of transporting a causal effect from a randomized trial to a specific target population. We present an inverse odds weighting approach that can easily operationalize transportability. We derive these weights in closed form and illustrate their use with a simple numerical example. We discuss how the conditions required for the identification of internally valid causal effects are translated to apply to the identification of externally valid causal effects.*	Westreich[18]